

Role of the DASD storage control in an Enterprise Systems Connection environment

by C. P. Grossman

This paper compares the Enterprise Systems Connection Architecture™ (ESCON™), with its use of fiber optic cables, to the parallel channel architecture introduced with System/360™. It also describes many of the reasons for the introduction of ESCON. The ESCON implementation for the IBM 3990 Storage Control is described in some detail, including a description of nonsynchronous operation. The paper concludes with a discussion of some of the benefits of ESCON for a 3990 installation, performance considerations, and migration considerations.

How peripheral devices are connected to a central processor is often taken for granted. Much attention is focused on the design of the processor itself, and the input/output devices also receive much study. However, the ways in which these devices are connected to the host processor for transfer of data and control of their operation can have very significant ramifications on system performance, application availability, and effective management of the data processing complex.¹ IBM's large systems have employed an attachment architecture using parallel channels, a set of channel command words using the Start IO instruction and subsequently the Start Subchannel instruction, and an addressing scheme based on the channel subsystem hardware of the System/360*, System/370*, and System/390* processors. This architecture has performed well for more than a quarter of a century and evolved over time to accommodate an extensive range of

processors, applications, and peripheral devices. Current trends in the data processing industry have shown that this architecture now requires very fundamental changes to continue to perform well for the large systems of the 1990s. Thus, the Enterprise Systems Connection (ESCON*) Architecture* has been introduced to meet the changing needs of current data processing installations.

The ESCON architecture was developed in recognition of a number of operational requirements of today's data processing installations. These requirements include relieving the constraint that limits the physical location of I/O devices to within 400 feet of the processor, providing a lighter-weight and less bulky replacement for the parallel bus and tag cables, and providing higher data transfer rates. Today's installations also need significant new capabilities in the area of configuration management (including controlled configuration changes) and substantial improvements in total systems availability.

To accommodate these requirements, not only is a fresh architectural approach needed as to how large processors communicate with and are con-

©Copyright 1992 by International Business Machines Corporation. Copying in printed form for private use is permitted without payment of royalty provided that (1) each reproduction is done without alteration and (2) the *Journal* reference and IBM copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied or distributed royalty free without further permission by computer-based and other information-service systems. Permission to *republish* any other portion of this paper must be obtained from the Editor.

nected to I/O devices, but so is the introduction of new cable technology for data transmission and special products for managing the resulting network. Changing the channel transmission medium relieves one set of concerns by providing higher data transfer rates, increased distances between channels and control units, and reduced weight and bulk of cable connections, but does not by itself address the other, potentially more important, items without the architectural changes.

With the introduction of the ESCON architecture, significant changes have been made to the IBM 3990 Storage Control and its supporting software in order for it to function in the new environment. (Note that even though this paper focuses on the IBM 3990 Storage Control, most of the discussion also applies to the IBM 9343 Storage Controller.) The areas of major change include:

- Different channel topology and resultant changes in the way communications between channels and devices are handled (by "channel topology" we mean the ways in which channels and devices are cabled together)
- Extension of our system capabilities for providing high availability and continuous operations
- Extended Count Key Data (ECKD*) architecture and nonsynchronous operation
- Configuration awareness across the entire ESCON installation and associated problem analysis
- Protection of current hardware and software structures as much as possible
- Provision of a variety of migration paths

The ESCON architecture

In many ways, the ESCON architecture builds upon the parallel channel architecture we have used for controlling the data flow between processors and peripheral devices, and in other ways it introduces substantial changes to the means of interconnecting devices. ESCON allows for more flexibility in the placement of peripheral devices, up to nine kilometers (and even more for some devices) from the central processor in comparison to the 400 feet allowed in the parallel configuration. Use of fiber optic cables in ESCON facilitates transmission of data at substantially higher rates than are possible with the copper wire cables used in a parallel configuration. The architecture provides new capabilities to modify and add to a configuration without disruption of cur-

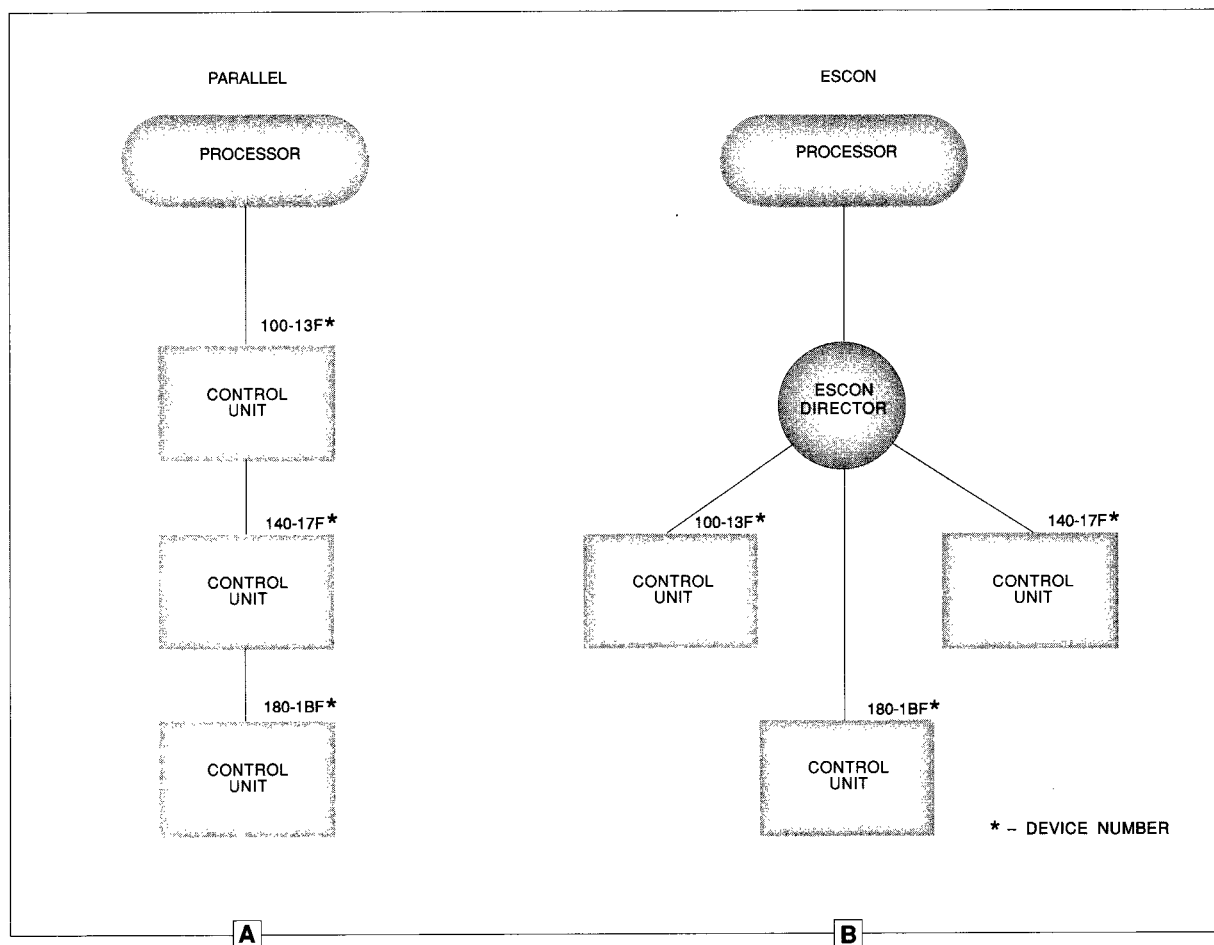
rent users of the system. Its centralized structure, when using the directors and the ESCON Manager*, allows for much greater levels of systems management of configurations. In the course of this paper, we investigate the ways in which ESCON implements these capabilities.

The most apparent change between the parallel I/O architecture that has been used for more than 25 years and the newer ESCON architecture is the replacement of pairs of copper bus and tag cables with fiber optic cables. However, the architectural changes extend far beyond the mere replacement of one medium for data transmission with another. With the introduction of the ESCON Director*, essentially a very high-performance routing device, the topology of channel and control unit connection has also totally changed. Implementation of this topology has necessitated a new way of addressing components in an I/O system—link- and device-level protocols and frames.

Bus and tag topology. The bus and tag (or parallel I/O) architecture uses pairs of cables to allow communication between processor channels and I/O devices. Data and instructions are sent on the bus cable, and control information indicating the nature of the transfer is sent on the tag cable. Each cable is comprised of many individual copper wires. Data are transmitted on a parallel channel parallel-by-byte—one wire in the cable is used for each bit in the byte, plus a parity bit. Transfer is essentially one byte at a time. Very precise synchronization of the transmitted signals is required for proper operation and is one of the reasons that the maximum cable length allowed for parallel connections is 400 feet.

The bus and tag architecture allows multiple control units to be connected in series on a given channel path; this type of connection is called multidropping (or daisy-chaining) control units. Such a channel configuration is shown in Figure 1, part A. Attaching multiple control units to a channel is done because a configuration will typically have many more control units than channels and multiple channel connections per control unit. Often, the performance characteristics of I/O devices are such that many devices may share a channel with little or moderate impact on overall system performance, the impact varying with installation characteristics. Furthermore, having multiple paths to a set of devices results in significant benefits from load balancing and im-

Figure 1 Channel configurations with multiple control units per channel



proved performance. Another factor leading to having many control units per channel is the dominance of multiple processor installations, with direct access storage devices (DASD) shared among them. Often in such situations, devices are physically cabled to all processors but actively used by only one or two of them. The remaining attachments are provided to accommodate either occasional use or use in backup mode.

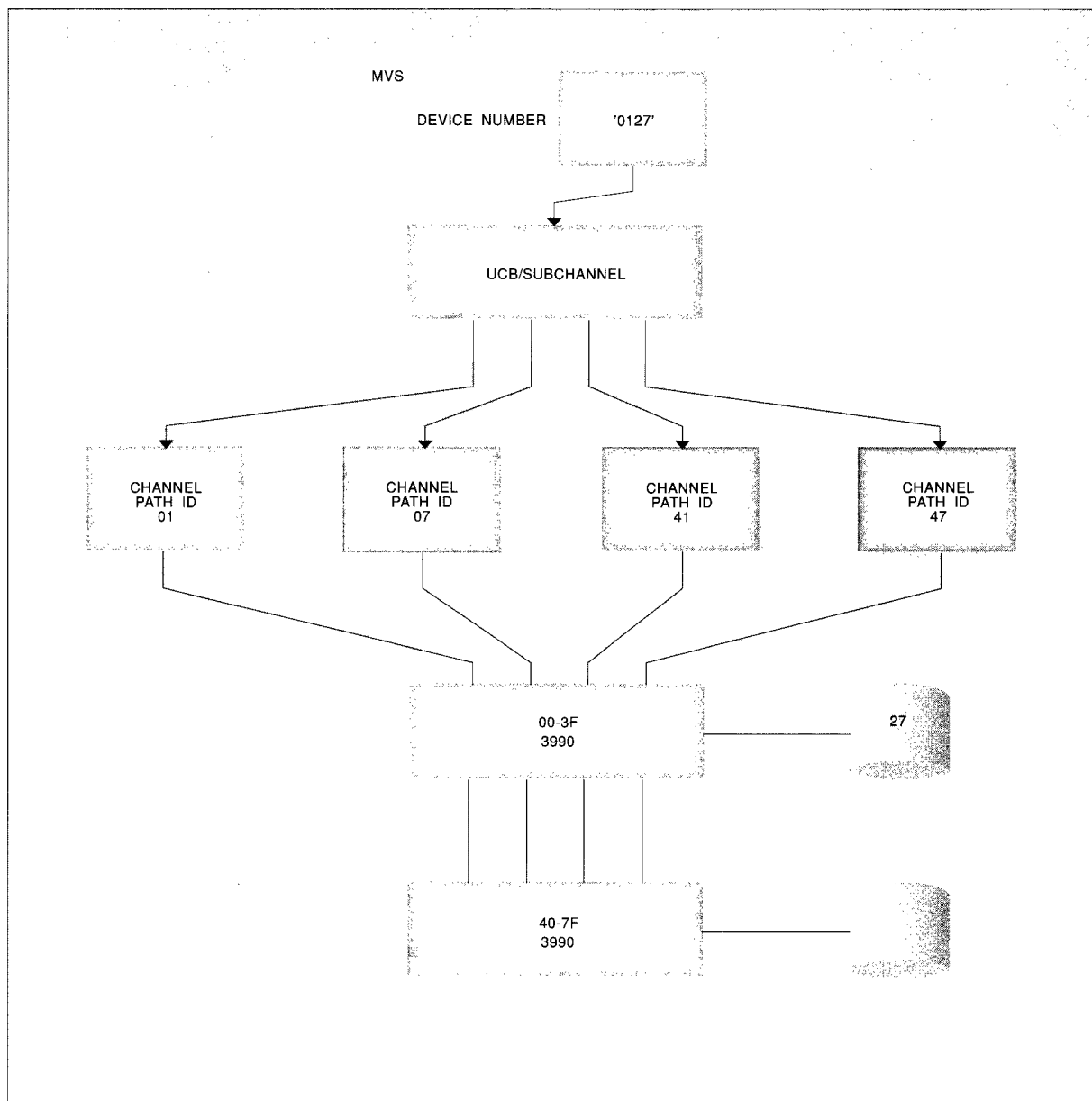
Parallel channel communication with I/O devices. The host operating system uses device numbers to designate specific I/O devices. Each device number is associated with the specific subchannel describing the device. In turn, this subchannel is translated into one of a specific set of connections from the system image to the device. (The exact number and nature of these connections is device-

dependent. DASD will allow up to four paths.) A device address is sent down the specific channel path and selected by the control unit that manages the required device as can be seen in Figure 2. Notice that the control unit is set up to recognize a given range of addresses, and different control units on the same parallel channel will have different address ranges. This mechanism selects the proper device on a channel with multiple control units.^{2,3}

After this initial selection, the control unit, device, and channel prepare for and complete the I/O operation through a series of transfers of control information, instructions, and data.⁴

Overview of ESCON components. The ESCON architecture is implemented through a variety of

Figure 2 Parallel I/O device selection



different products and components that work together to provide the system enhancements described above.

The 3990 Storage Control equipped with ESCON adapters operates in concert with all of the other components of ESCON. These products include new levels of the operating systems, the new

ESCON channels on selected IBM ES/3090* and all ES/9000* processors, the ESCON Directors, the ESCON Manager program product, and conversion devices.⁵

The ESCON Director provides the dynamic switching function that is the heart of ESCON. It is an extremely fast nonblocking switch. (A switch

is nonblocking when the existence of a connection between any two ports has no impact on the ability of any other pairs of ports to establish a connection.) Every channel or control unit in an ESCON environment may attach to the director through one of its ports. When, for example, a channel wishes to communicate with a control unit, it sends addressing information to the director that enables the director to establish the port connection allowing the I/O operation to complete. These connections are established and terminated as required for each I/O operation. This switching function, together with the director's capability of responding to commands to change its port configuration, and the ability to plug a fiber optic jumper cable into a director port without disrupting other I/O operations, is what makes the operational enhancements of the ESCON architecture possible.

IBM processors with ESCON channels provide not only the new fiber optic channels but also the channel subsystem logic to manage the new configurations. Functions to be described in more detail later include initialization of the channel paths, recognition of the attached control units and directors, selection of path components to be used to complete an I/O operation, and recognition of error conditions and reporting them to the system control program.

The ESCON Manager is used to provide cross-processor information and control over an ESCON installation. Typically, it will run on each system image using ESCON within the installation. Because of the dynamic configuration potential of ESCON and the extended distances over which ESCON components may be installed, the ESCON Manager centralizes system monitoring and change.

Converter devices are available to allow non-ESCON control units to attach to ESCON channels and to allow ESCON 3990 Storage Controls to attach to parallel channels. These devices are described in more detail later.

ESCON topology. The ESCON architecture supports a switched star topology when used with the ESCON Director. Theoretically, any unit attached to the director may communicate with any other unit attached to the same director. In practice, limitations are imposed by the host software, the channel subsystem definition (the I/O configura-

tion program, or IOCP), and the specific characteristics of the control unit. It is of utmost significance that a redefinition of a particular channel configuration or change of the definition of the allowed paths to a control unit may be accomplished *without* physically recabling any devices or channels. This key feature of the ESCON architecture permits the configuration flexibility and nondisruptive modification of the I/O configuration (in conjunction with either a predefined system or the use of the dynamic configuration capability of Multiple Virtual Storage/Enterprise Systems Architecture Version 4, or MVS/ESA* 4.2). See Figure 3.

In a parallel configuration, the possible channel paths from a processor to a device are established at installation time, and the IOCP reflects the way in which the cables are physically installed. By contrast, in an ESCON configuration using the ESCON Director, logical paths are defined in the IOCP and established at hardware initialization. The ESCON Director provides the mechanism for switching from one to another. Figure 4, part B shows how four processors may be defined so that each has four paths to the 3990, and all four share the same four links from the 3990 to the director. This ESCON configuration provides equivalent connectivity to the parallel configuration shown in part A of Figure 4.

The components of a logical path are electronically connected for the duration of an I/O operation. Upon completion of one I/O operation, these components may be dynamically connected in a different logical path for another I/O operation. So, in Figure 4, for example, one I/O operation could use the logical path from processor X, channel 1, to interface A on the 3990, and a subsequent I/O operation from processor Y could use the same link from the director to interface A on the 3990.

Figure 4 shows the fundamental difference between the parallel channel topology and the ESCON topology. In a parallel configuration, the connections are static and define the possible paths. In ESCON, connections are dynamic, and fewer links from the storage control to the director are required to provide equivalent levels of performance. The 3990 is capable of sustaining four concurrent data transfer operations on the upper interface, so no more than four paths from the 3990 to the channels could be busy with data

Figure 3 The new ESCON topology

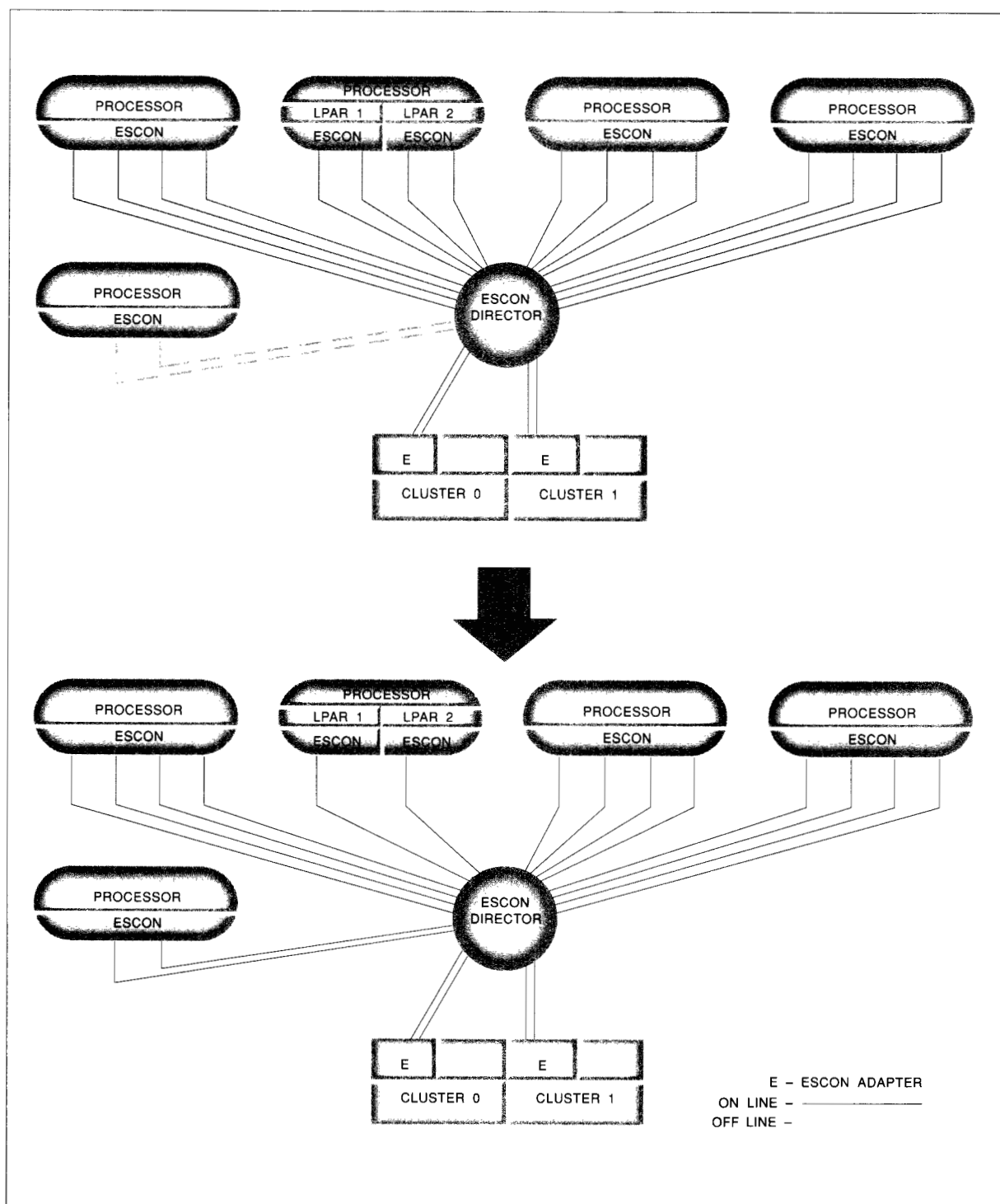


Figure 4 IBM 3990 with 16 parallel paths and an ESCON 3990 with equivalent pathing using the ESCON Director

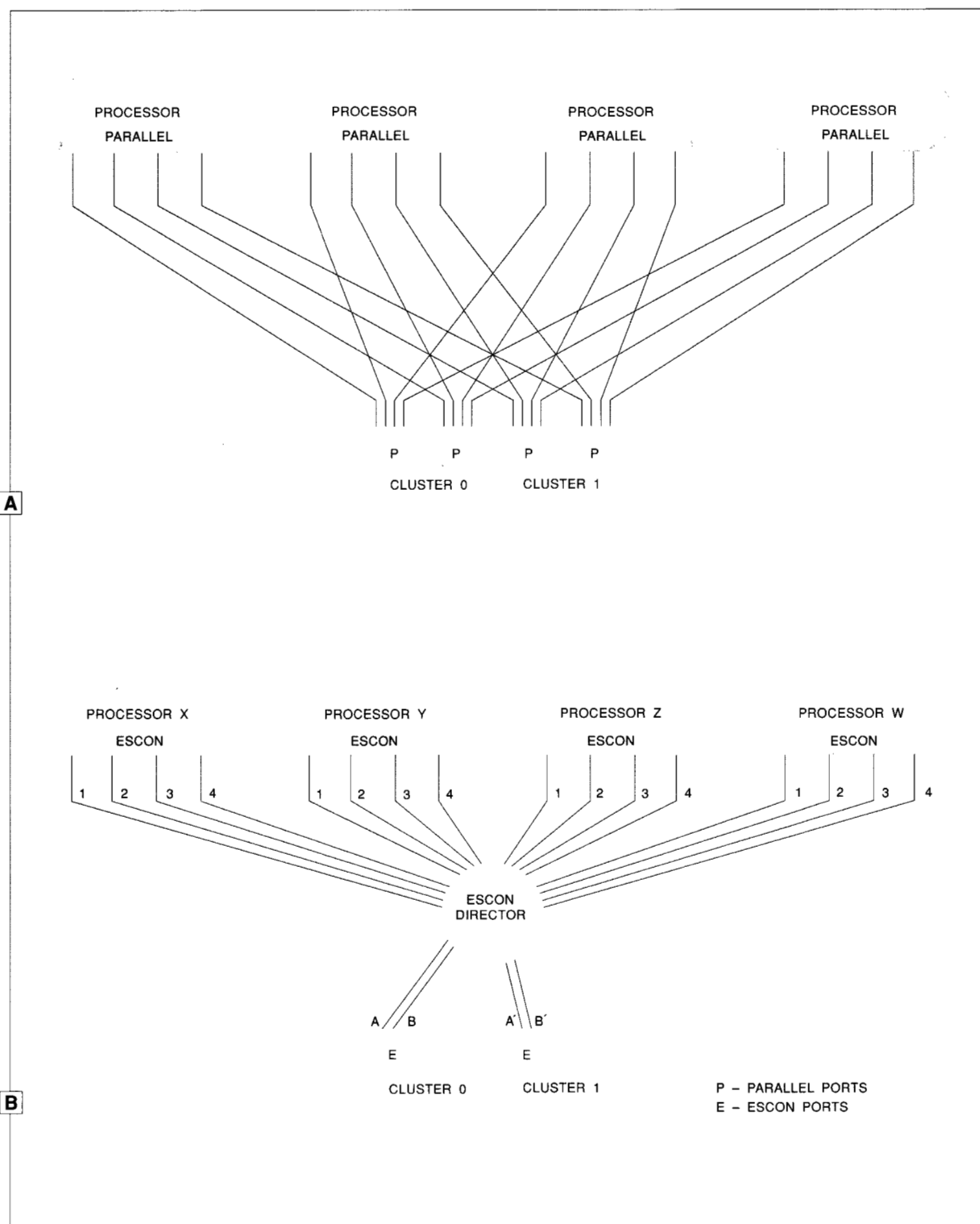
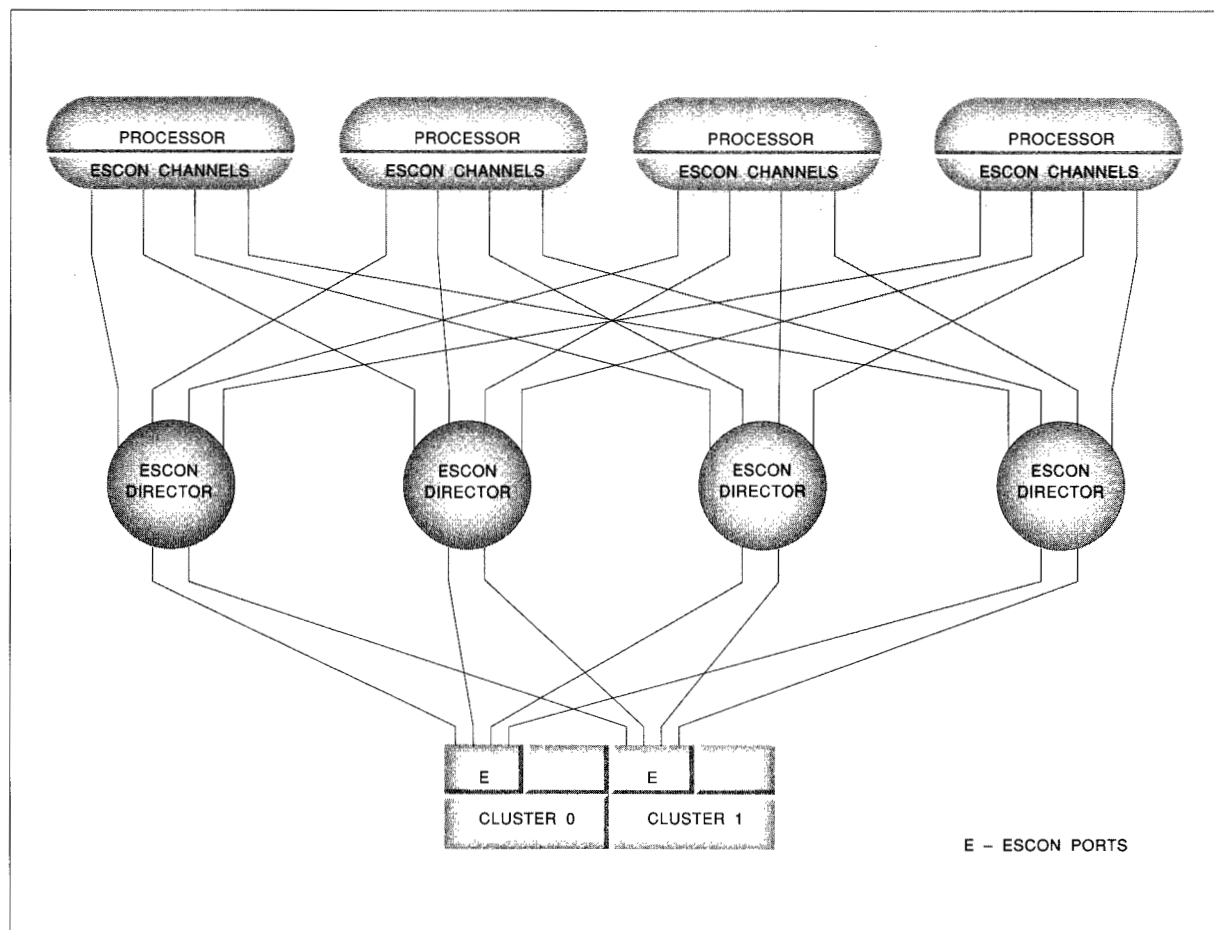


Figure 5 Configuration of an ESCON 3990 with four ESCON Directors for maximum availability



transfer. Because of the dynamic nature of ESCON, no more than four paths are needed from the 3990 for performance reasons. However, availability is also a consideration. For availability reasons, a configuration such as that shown in Figure 5 may be desirable. In such a configuration, eight 3990 ESCON ports (four per cluster) are used to attach to four ESCON Directors. The overall performance impact of a failure is minimized by distributing the paths across both clusters and all four directors. (In the 3990, a cluster consists of two storage paths and the associated hardware. There are two clusters, each on separate power and service boundaries so that one of them may be removed from use for servicing without causing loss of access to data.) This is an example of the highest levels of performance available in case

of a failure of one of the directors. Many other configurations providing intermediate levels of performance in such situations are also possible. It is, in fact, one of the strengths of ESCON: an installation has many more options for tailoring the configuration more closely to its particular combination of performance, availability, and operations requirements.^{6,7}

Because the ESCON topology is essentially point to point, only one physical control unit may be directly attached to a port. When it is desirable to have several control units use the same ESCON channel link, the ESCON Director must be used as shown in Figure 1, part B. This configuration provides an equivalent attachment to the parallel configuration shown in part A.

The switched point-to-point configurations discussed so far will probably be the prevalent ESCON configurations used with 3990s. There are two other possibilities. One of these is a point-to-point connection between a processor channel and a control unit, which does not use the ESCON Director. This connection would be used primarily for a smaller configuration that does not need the flexibility or change capabilities of a full ESCON configuration.

Another possibility arises when the 3990 is operated in DLS (device-level selection) mode instead of DLSE (device-level selection enhanced) mode. All of the configurations described so far have been DLSE mode 3990s, with four possible paths from the storage control to a system image. In DLSE mode, the two storage paths within one cluster of the 3990 comprise one multipath storage director. DLSE mode is used with either the 3390 DASD or with 3380 J or K DASD operating in four-path mode. Either of these DASD types may take full advantage of the four-way dynamic path reconnect capability of IBM Extended Architecture (XA) or Enterprise Systems Architecture (ESA) operating systems.

3990 DLS mode is used with the 3380 Standard, D, or E DASD, all of which operate in two-path mode. DLS may also be used with 3380 J and K DASD operating in two-path mode and using two-way dynamic path reconnect. The 3380 D, E, J, and K are supported with the 3990 ESCON adapters. In DLS mode, each of the storage paths within the 3990 is defined as a separate single-path storage director. Two of them, one from each of the clusters of the 3990, are paired together to define a two-path DASD subsystem. Thus, in DLS mode, a 3990 will consist of two logical DASD subsystems, each with its own distinct set of devices.⁸

In order to accommodate a DLS mode 3990, we make use of the concept of a logical control unit. The ESCON architecture allows up to 16 logical control units within a given physical control unit. The logical paths from processors to the physical control unit may be shared among the logical control units. The 3990 will allow for two logical control units. Logical DASD subsystem 1, consisting of single-path storage directors 0 and 2, corresponds to one ESCON logical control unit, and logical DASD subsystem 2, consisting of single-path storage directors 1 and 3, corresponds to the

other. These logical control units are defined in the IOCP.

The possible types of 3990 ESCON configurations are shown in Figure 6.

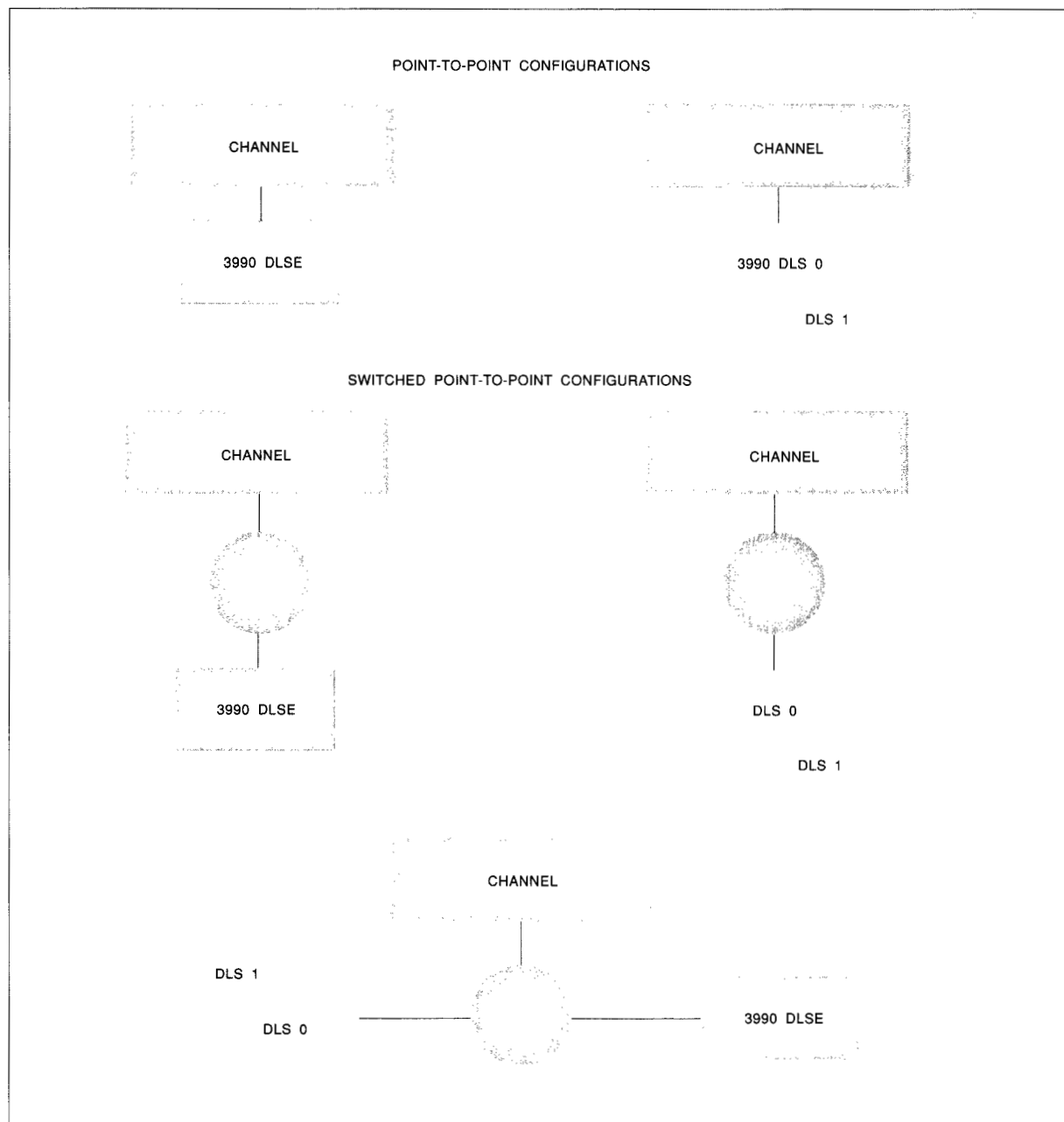
ESCON I/O operations. Protocols and transmission are totally different for an ESCON link. On an ESCON link, data are transferred serially-by-bit, and the data (or control information or instruction) are transmitted in frames. The size and type of frame vary with the intent of the transfer operation.

Included in every frame is a source and a destination address. Because a connection only exists for the duration of an I/O operation in a switched point-to-point configuration, a different technique must be used to identify the components of a logical path for a given operation—we can no longer rely on the physical connections to define the path. From a given channel to a device, the path is uniquely determined by the port address on the director, the logical control unit number (if used), and the device address within the control unit (for the 3990, a value from 00 to 63—hexadecimal 3F). This addressing information is included in every exchange between a channel and a device. See Figure 7.⁹

There are two types of frames: link level and device level. Link-level protocols and frames deal with housekeeping functions: status of links and ports, initialization, management of logical paths, link error reporting, etc. In short, the link-level protocols deal with the physical aspects of the configuration. For an I/O operation, link-level addressing determines the physical path—the director port address and the logical control unit number. Device-level addressing will specify the target device for the operation. Device-level frames provide the functions involved in actual data transfer (the channel command words), functional commands to control units, and reporting of control unit status (that is, sense information regarding error conditions or unusual conditions within the control unit and its devices).

Protocols and frames. The handling of the link- and device-level protocols as well as the construction and interpretation of the frames is isolated to the ESCON adapters of the 3990. Likewise, the bulk of similar activity in the processor is handled by the channel subsystem hardware. Much

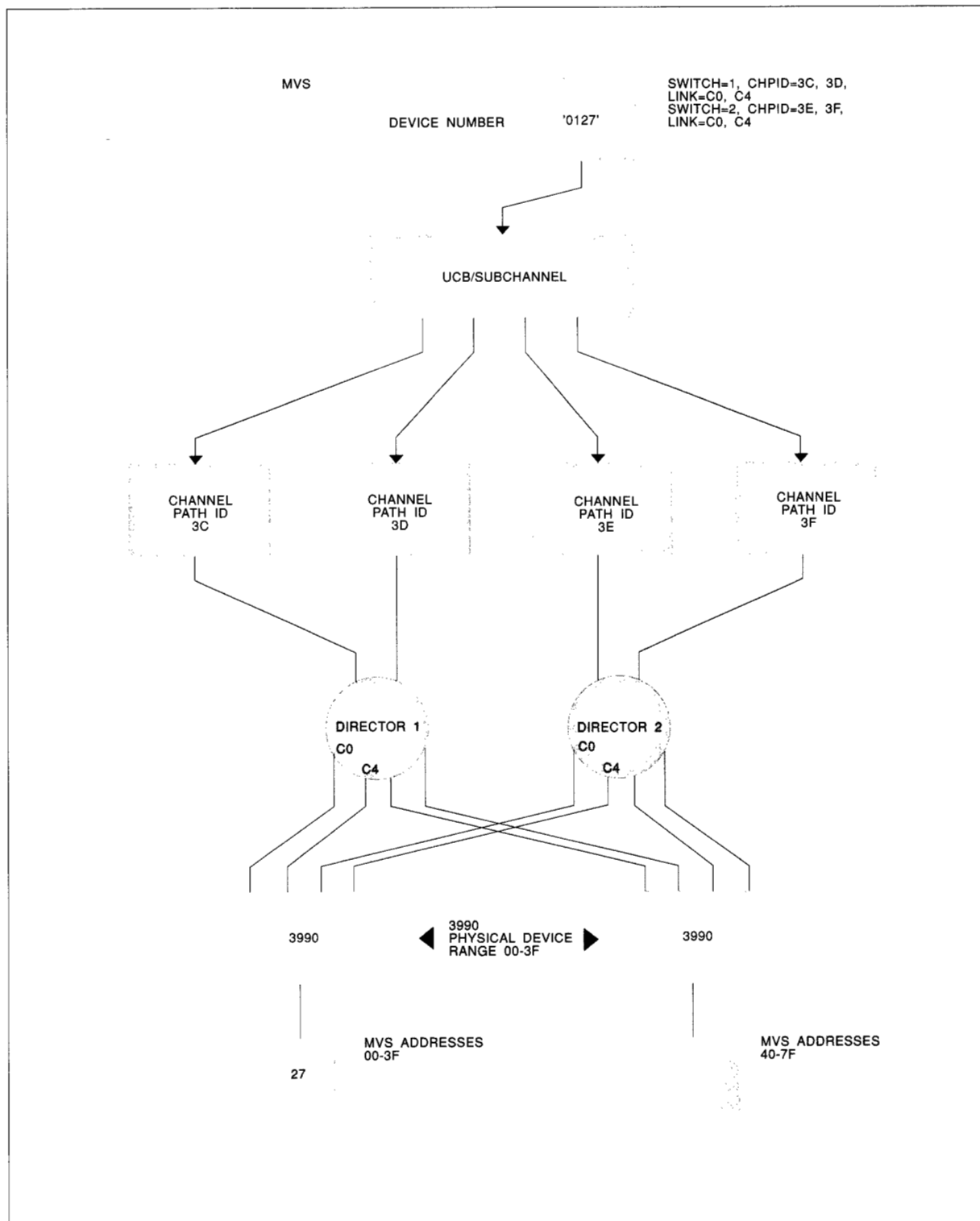
Figure 6 Possible ESCON configurations with a 3990



of the 3990 licensed internal code and the host IOS programs are relatively untouched by the ESCON changes. In particular, the segmenting of a record of data into several frames for transmission across the fiber optic links is totally transparent to

I/O buffering in the host and storage of blocks of data on the DASD. For a write operation, the re-assembly of the record from the frames is done in the automatic data transfer (ADT) buffer before it is written to DASD as a block. The ADT buffer is a

Figure 7 ESCON device selection



64K-byte buffer that resides in the 3990 storage path. Each storage path has one buffer, and no buffers are shared. The ADT buffer is also used for two-level error correction code (ECC).⁶

ESCON configuration management

The rapidly increasing complexity of large data processing installations, with a very large number of devices, several processors, and extensive use of the Processor Resource/Systems Manager* (PR/SM*) logical partitioning capability, make centralized and more automated control and monitoring of configurations mandatory. (PR/SM is a processor feature that allows hardware resources, such as CPU cycles, processor storage, and channels to be logically partitioned so that independent system control programs may execute on each separate partition independently from any of the other partitions.) By design, the ESCON architecture provides the structure to effect this control.

Configuration awareness. A major component of this structure is providing information about the ESCON configuration. The primary sources of information about the configuration are the IOCP and the ESCON Director (ESCD) configuration definition (whether defined at the ESCD console or via the ESCON Manager). Virtually all of the remaining descriptive information is derived from these two.¹⁰⁻¹²

When the processor is powered on (power-on-reset), the I/O Configuration Data Set (IOCDS—the output of the IOCP) is used to create the hardware system area (HSA) used by the channel subsystem. When the channel subsystem is initialized, it will establish logical paths to all of the ESCON control units. In the process, all of the required channel path and addressing information will be tabulated within the 3990. At IPL (initial program load) time, output from either the MVS configuration program (MVSCP) or the HCPRIO program (for Virtual Machine/Enterprise Systems Architecture, or VM/ESA*, systems) is used to establish the I/O device control block structure. Also at IPL time, the ESCON Manager will use the HSA information and connection information obtained from all of the attached ESCON Directors to build its own set of control blocks describing the ESCON configuration.¹³

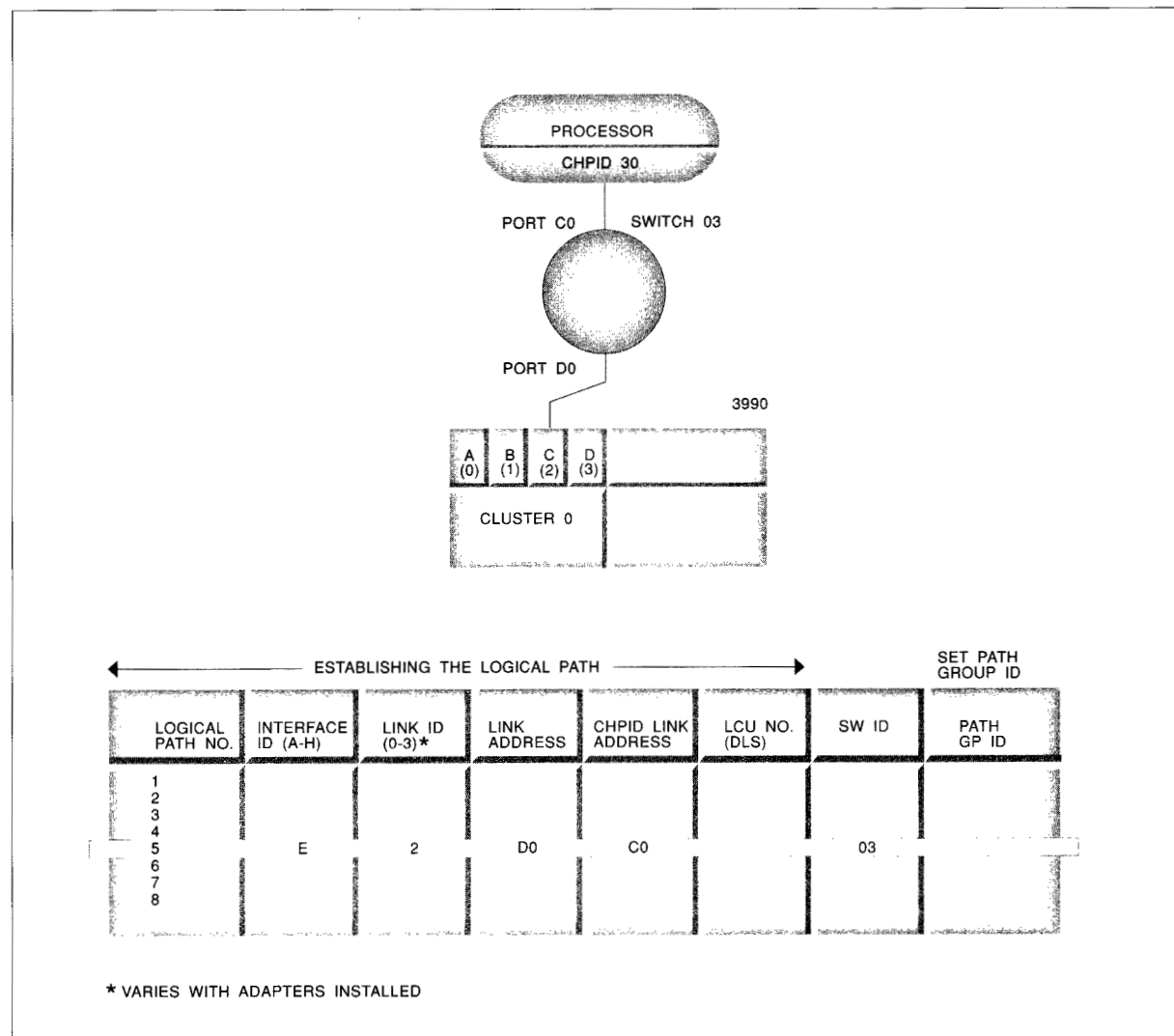
The 3990 vital products data (VPD) are the source for all of the configuration information provided

to the 3990. All of this information is provided at installation time, at configuration change time, or when logical paths are established. Much of the information is entered by means of the 3990 Support Facility, an integrated microprocessor used for both configuration and maintenance. Data include cache size, mode of operation (DLS or DLSE), and other 3990-specific items. For parallel channel attachment, data are used to specify the channel speed and address ranges used for each channel attached.¹⁴

For ESCON ports used with ESCON channels, the VPD requires no more input than an indication that ESCON adapters are installed. All of the remaining information required by the 3990 is generated at initialization time, as described below. In the case of the 3990, this means that host ESCON channel configurations can be changed in a variety of ways without requiring any action at the 3990 Support Facility. Consequently, addition or deletion of processor paths can be accomplished without disturbing the workload or performance levels of the operational processors. These configuration changes can also be automated with a program product such as NetView* working in conjunction with the ESCON Manager, and executed much more efficiently. This type of centralized control is particularly desirable because the ESCON fiber optic links may now greatly extend the distances between channels and I/O devices. Effecting configuration changes at a storage control located a kilometer or two away from the master console is just not practicable.

Initialization. At the time a processor image is powered on or IMLed (IML—initial microprogram load), each ESCON channel executes an initialization procedure which, among other things, determines the link address for the channel (i.e., the ESCON Director port address). The 3990 sends the “acquire link address” request to the director to determine its own port address. The channel uses the IOCP information to issue commands to establish logical paths to all control units defined on that channel. The 3990 then builds the definition of the logical path consisting of the logical path number; the link identifier, or ID (the physical port number on the ESCON adapter); the channel link address; and an assigned logical channel interface ID (A—H) in the 3990. This logical channel interface ID is no longer uniquely associated with the port. Two or more interface IDs could be associated with one physical port. Finally, the channel

Figure 8 ESCON logical path initialization for a 3990



and the 3990 exchange information to establish the maximum data transfer rate to be used between the channel and the control unit. This transfer rate is dependent upon channel characteristics and the control unit capabilities. See Figure 8 for a diagram of the initialization process.

Path management functions. The path management functions are critical to system operation. Because processors with and without ESCON channels may share the same 3990, it is vital that

these functions operate in the same way for both parallel and ESCON paths.

- Path groups are created at initialization of the paths from a given processor image to a 3990. Typically, two or four paths will be included in a path group, depending on the DASD capability for two-path or four-path operation.
- Dynamic path reconnect allows an I/O operation initiated on one path to complete on any other

path within the same path group. This function provides availability and performance benefits.

- Reserve/release is a mechanism used to preserve data integrity in a shared DASD environment. When a reserve is issued for a device by one processor, the storage control will only allow the paths in the path group of that processor to access the designated device until the same processor issues a release.
- Channel selection algorithms are used to balance the I/O load across all of the paths to the device.
- Alternate path selection is a mechanism for restarting an I/O operation on another path in the path group when the initial path has a busy component.^{2,15}

Externally, all of these functions operate in the same way as they do in a parallel configuration. One distinction, as we have seen, is that instead of defining a path using the physical interface (A—H), the ESCON 3990 defines a logical path using the port identifier and the channel port address on the director. All of this is done within the 3990 ESCON adapter. The ESCON adapter associates the port identifier and the port address with an interface identifier used by the storage path. The result is that even in an intermixed parallel and ESCON 3990, these functions operate as usual.

In an ESCON environment, an alternate path is used for the same reason as in the parallel environment: some component of the I/O path to the device is found to be busy. Now we may see a port busy condition in addition to the other possible busy conditions. The effect is essentially the same as encountering a control unit busy condition.

As has always been the case, all of the logical paths within a path group (that is, all of the logical channel paths from a given system image to a given 3990) must be of the same type. All of the possible paths for dynamic reconnect or alternate path selection must be the same type as the initial path selected for the I/O operation. From the perspective of the processor, three channel types attach to a 3990: block multiplexer, ESCON, and ESCON configured for an IBM 9034 Converter. These three channels are distinct, with unique characteristics that preclude connection within the same path group. (Note that selected ES/9000 processors allow intermixing of block multiplexer

and ESCON channels configured for 9034 Converters.¹⁵)

For similar reasons, the 3990 also recognizes three channel connection types: block multiplexer, ESCON, and block multiplexer operating with an IBM 9035 Converter. The 3990 does not allow intermixing of these channel types within a path group. See the last section for more detail on the characteristics of the converters.

Nondisruptive change in conjunction with dynamic I/O configuration. Large data processing installations need the ability to reconfigure hardware without disrupting users of the system, and preferably with minimal impact on performance. Reconfiguration includes not only adding new devices, but also rearranging devices for load balancing or to accommodate new applications or new processors. Until recently, making such changes with the least impact possible was difficult and required anticipating future hardware increments. In order to later add a new DASD subsystem, for example, the installation had to anticipate this addition by including a description of the new devices in both the IOCP and the MVS or VM system. Otherwise, a power-on-reset of the processor and an IPL of the host software would be required to activate the new IOCP and the software description of the new devices. Even if this work had been done in advance, the actual hardware installation would be done one path at a time, with a potential impact on performance of all applications using other devices on the same path.

The ESCON architecture and dynamic I/O configuration, a capability of MVS/ESA 4.2 in conjunction with the ES/9000 processors, virtually eliminates these exposures. A new storage control may be physically installed by attaching its fiber optic links to unused ports on the director. This installation has no adverse impact on any other users of the director. There is no disruption on any of the ESCON channels. The MVS/ESA 4.2 hardware configuration definition and dynamic I/O configuration may be used to define the new devices to the system and activate them without causing a power-on-reset or IPL of the host. To go even one step further, the description of a device to the host may be modified without causing an IPL. Possibilities include changing its physical link connections or its allowed channel connections through the director, or changing the eso-

ECKD architecture and nonsynchronous operation

The count-key-data (CKD) channel programs that have been developed over the last quarter century rely very heavily on both the geometry of the device and the synchronous nature of DASD operation. Over the years, many channel programs have been very precisely tuned to take advantage of device characteristics to optimize performance. This optimization has come at the expense of major rework of the channel programs when a new device is introduced. One of the key elements is the gap between adjacent fields on the track. For illustration, suppose that a channel program consisting of a search ID command followed by a read command is being processed (Figure 9). During the time the read-write head passes over the gap between the count field and the data field, the storage control sends the channel information about the field just read, receives the read channel command from the channel, and completes the necessary internal processing so

WITHIN A CHANNEL PROGRAM, THE NEXT CHANNEL
COMMAND MUST BE RECEIVED BEFORE THE
INTER-RECORD GAP PASSES THE READ-WRITE HEAD

SEARCH ID

"COMPLETE" ←

READ RECORD CMD → GAP

"BEGIN
DATA TRANSFER" ←

RECORD

A major component of DASD design has been optimizing the gap size with the internal operating characteristics of the storage control and the device, the device linear bit density, and the 400-foot distance allowed for parallel channels. In order to allow DASD operations beyond 400 feet, a new mode of operating between a storage control and a channel had to be introduced. Because the distances needed could vary so much, just increasing the gap size is not desirable. The other alternative is to change the processing technique so that the channel is not required to be so closely synchronized with the device—the channel is no longer required to send the next command within the interrecord gap. This alternative is accomplished through the introduction of nonsynchronous operation, facilitated by the Extended Count Key Data (ECKD) architecture and the use of a buffer in the storage control.

One of the key features of the ECKD architecture is its deliberate definitions of records, tracks, and

The ECKD command set is built on the CKD command set.

cylinders that are divorced from a particular disk device implementation. For example, a cylinder is a collection of related tracks, with no implied physical relationship. The only constraint is that each cylinder must contain the same number of tracks. A track is some portion of a 360-degree revolution of the disk. Record IDs may be any five-digit string, as long as they are unique within a track. Defining terms in this way is intended to focus attention on the fact that old relationships are no longer valid and may not be assumed. In addition, this approach has the advantage of ensuring that the channel programs are not device-dependent and also of encouraging simpler coding techniques.

The ECKD command set is built on the CKD command set and relies very heavily on two channel commands, LOCATE RECORD and DEFINE EXTENT. These commands were initially used for the IBM 3880 Speed Matching Buffer and 3880 support for fixed-block architecture devices. They have been subsequently modified for 3880 and 3990 cache functions as well.¹⁸ DEFINE EXTENT is used to establish the permitted range of tracks accessed by the channel program. The channel program will not be allowed to access tracks outside this range. LOCATE RECORD is used to tell the storage control in advance what records are to be accessed and the type of operation. This advance information allows the storage control to optimize its internal operation for the impending request. (In fact, this is of some benefit to performance even in a synchronous environment.)¹⁸

In nonsynchronous mode, cache read hits and DASD fast write hits are processed exactly as they are in synchronous mode. For a read hit, the record is read from the cache and transferred at channel speed to the host. For a DASD fast write hit, the record is written to the cache, and the

nonvolatile storage (NVS) and channel end, device end are issued. The transfer is at channel speed.

A CKD channel program *will* execute properly in a nonsynchronous storage control. In general, a CKD read channel program should execute with little or no change in performance. However, for CKD write channel programs or channel programs with seek head commands, performance may not be as expected since there will be a rotation for each write or seek head. For example, in a CKD write channel program, a search command is used to locate the record to be updated. The track is staged into the ADT buffer to complete this search operation. Now, when the write command arrives, the read-write head is physically beyond the beginning of the data area. The data transfer must wait until the record comes under the read-write head again. In the case of an ECKD channel program, this sequence is not required—the storage control notifies the channel when the read-write head approaches the desired record, and the channel sends the write command and data to the buffer, from which it is written to the device. An ECKD channel program will execute in synchronous mode. (In fact ECKD channel programs have been executing on the 3990 in synchronous mode since it was introduced in 1988.) Thus, there is no implied relationship between the command set being executed and the mode of operation of the storage control. The only connection is that ECKD was specifically designed to optimize the performance of a nonsynchronous storage control.

ECKD architecture and access methods. The introduction of nonsynchronous DASD operation has necessitated changes in the structure of channel programs, compared to what was done for earlier DASD storage controls. In anticipation of the need for restructured channel programs with ESCON nonsynchronous operation, ECKD was introduced with the 3990. As we have seen, this architecture relies on the DEFINE EXTENT and LOCATE RECORD channel commands as well as on a set of rules for writing channel programs to operate properly in this environment. The standard IBM access methods conform to the ECKD architecture. Well-formed CKD channel programs that include a search ID for orientation and consist only of read operations, or have only one update write, or one or more format writes, will be detected by an EXCP (EXECUTE CHANNEL PROGRAM) scan routine and have the appropriate DEFINE EXTENT and LOCATE RECORD commands prefixed to

the channel program. With this modification, they will function in the 3990 as ECKD channel programs. CKD channel programs that do not qualify (typically they have multiple writes or key operations) will be executed as CKD channel programs with no advance indication to the storage control of the operation to be performed. In this case, an impact on performance is likely, due both to the fact that there will be an additional disk rotation before the write is complete and to possible additional anticipatory staging of data. Because the 3990 has received no advance information regarding the type of operation (read or write), it assumes the data transfer will be a read and begins to read data from the device into the ADT buffer when the search completes. When a write channel command is received, the data in the buffer are discarded, and channel command retry is invoked to redrive the write. Since the channel program might not use these data, the stage may be ineffective and needlessly use a storage path.

The important thing to note is that in many cases the impact on performance of operating CKD channel programs nonsynchronously is minimal. A very large number of such channel programs will be converted by the scan routine. These programs should perform reasonably well—there may be some extraneous staging because the 3990 may stage more blocks than required. Write operations and key searches are typically a small portion of the total number of I/O operations. Performance-sensitive applications that use CKD channel programs will need to be analyzed for use of channel programs that will not be converted by the scan routine. These applications will include channel programs that have:

- Key orientation
- Multiple update writes
- Imbedded seeks or searches, or both
- Program-controlled interrupt (can be used to modify channel programs while they are executing—a violation of the ECKD architecture)

These channel programs will execute as CKD channel programs and should be evaluated for total impact on system performance. If they are not performance-sensitive or are not used with great frequency, there may be no need to convert them. The expense of converting them may outweigh the performance improvement gained. However, if they are used frequently or are response-time sensitive, it may be worthwhile to convert them.

ESCON performance considerations

ESCON link and device protocols and nonsynchronous DASD operation introduce several new performance evaluation factors and modify others. These factors include:

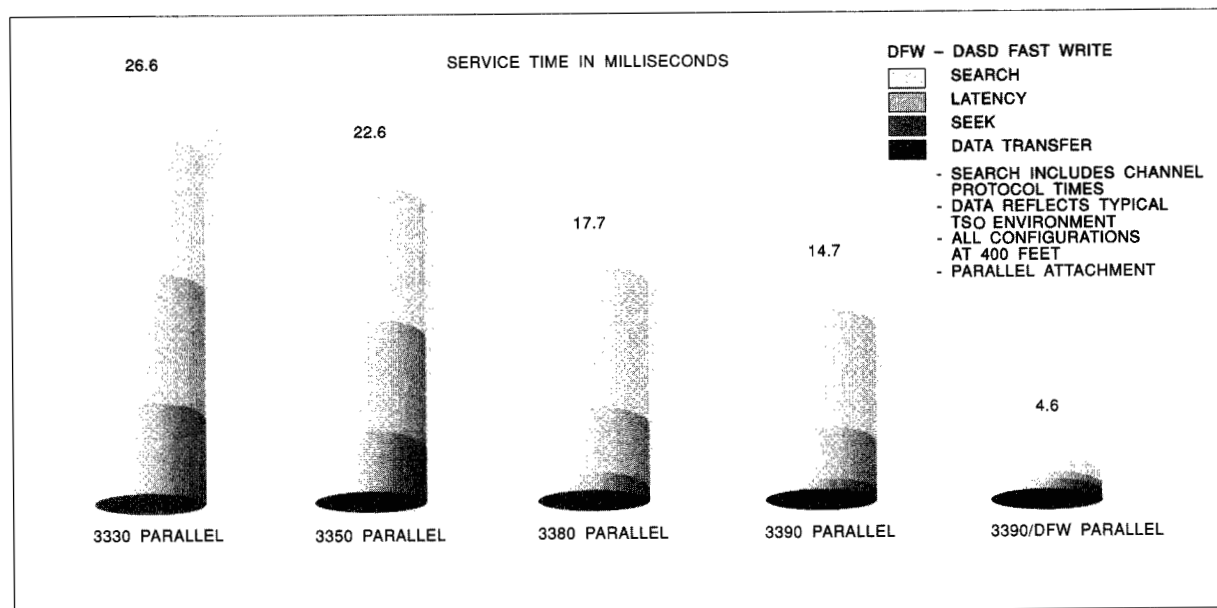
- Channel transfer rate
- Presence of link and device frames
- Channel turnarounds
- Distance
- Nonsynchronous algorithms
- CKD/ECKD channel programs
- Configuration
- DLS/DLSE operation
- Differences between cached and noncached operation

The breadth of change shows that our existing rules of thumb for DASD performance will be modified and supplemented for the ESCON environment. As we gain more experience with ESCON, both through the ESCON performance models and through user experience, we will need to challenge the rules of thumb we have used for years and modify them appropriately.

The most obvious of these factors affecting performance is the channel transfer rate. Since the time of the IBM 3330, we have seen an almost five-fold increase in DASD data transfer rate—from 0.886 megabytes (MB) per second to 4.2 MB per second for the 3390. Because data transfer was such a large component of DASD I/O activity for the 3330, introduction of the IBM 3350 with its 50 percent increase in data transfer rate resulted in a significant improvement in DASD performance.¹⁹ However, as the data transfer component (Figure 10) of an I/O operation decreased, the percentage improvement provided by an increased transfer rate has dropped. For a 3990 Model 3, introduction of the 4.5 MB per second data transfer rate can result in about a 5 to 10 percent improvement in performance, compared to the same configuration and workload measured on channels having a rate of 3 MB per second.

Similar considerations apply to the ESCON environment. Noncached data transfer occurs at device speed. Cache hits, whether read hits or DASD fast write hits, transfer at ESCON channel speeds of 10 or 17 megabytes per second, depending on the attached processor. So, the opportunity for an improvement in performance due to channel

Figure 10 Components of a DASD I/O operation, from the 3330 to the 3390 with DASD fast write



speed is thus limited to 3990 Model 3 cache hits. To the extent that applications have good caching characteristics, they may experience an improvement in performance compared to a similar parallel configuration (at equal channel distances up to 400 feet). In general, the response time of a 3990 Model 3 with ESCON paths of 400 feet will be comparable to or slightly better than that of a parallel 3990 Model 3 with the same DASD configuration and workload.

For either parallel or ESCON configurations, transfer of data and channel commands is only part of the traffic on the channel. In either case, special sequences of control information are sent back and forth between the channel and the control unit to prepare for an I/O operation. These channel turnarounds contribute to DASD subsystem performance. Although the number of turnarounds for ESCON configurations is substantially less than for parallel configurations, there could still be as many as ten or more turnarounds per read or write channel program, depending on the type of I/O operation.^{4,9}

These channel turnarounds are the reason ESCON performance varies with the channel distance. The turnarounds will contribute additional ser-

vice time on the ESCON channel. As channel distances become longer, this becomes a small but measurable component of DASD service time. Measurements have shown that ESCON 3990 response time could vary by approximately one to two milliseconds as the length of the fiber optic path extends from 400 feet to nine kilometers.⁶

The primary determining factors for fiber optic path lengths should be installation requirements for the physical location of system components and availability requirements to have paths follow different physical routes for protection from inadvertent link damage. However, considerations for economy of an installation and good general systems design both suggest keeping all paths as short as is reasonable.

ECKD and nonsynchronous operation of the storage control may also affect performance assumptions. Many channel programs directed at a 3990 are already adhering to the ECKD architecture. CKD channel programs should be converted to ECKD. As we have seen, CKD channel programs may impact performance in several ways: unnecessary staging of data and extra device rotations for writes.

The 3990 uses the automatic data transfer buffer and, for the Model 3, the cache, for nonsynchronous processing. For the 3990 Model 3, read and write hits are processed, as usual, directly from the cache. No channel disconnect is required because there is no need for device positioning. Thus, for hits, there is no rotational position sensing miss, and any transfer is at the full channel speed. (Rotational position sensing—RPS—is a capability of IBM DASD that allows the device to disconnect from the channel during the time that it takes for the read-write head to approach the desired record. The control unit notifies the channel when the read-write head gets close to the desired record, and the channel reconnects. RPS thus improves channel utilization. An RPS miss occurs when the device attempts to reconnect to the channel path after reaching the desired location on the track, but some portion of the path is not free to complete the operation.) For I/O activity to a 3990 Model 2, the storage control must acquire the full path from the device to the channel. Thus, RPS miss considerations are the same as for parallel channels.

In the parallel-channel environment, performance and availability have been major factors in configuring DASD subsystems. The 3990 Model 3s with 3390 DASD, using DASD fast write and four channel paths to each attached processor image, provide not only the best performance for an IBM DASD subsystem, but also excellent subsystem availability. (Dual copy may be used to further improve DASD availability.) Four-path operation reduces RPS miss very significantly, almost eliminating that component of an I/O operation. In a four-path configuration, the performance impact of losing one path (whether due to hardware failure or intensive use of one path) is significantly less.

The same four-path configuration with ESCON channels will likewise provide the best possible performance and availability. The only additional considerations from an availability standpoint are the choice of an ESCON Director configuration and the physical planning for security of the links.

Many installations today have either fewer than four paths from a given processor to a storage control or more than one storage control per channel path, or they have both. To the extent that these configurations provide satisfactory performance with parallel channels, they should be

used as a starting point for the ESCON configuration.

However, trying to take advantage of the higher channel rates to reduce the number of channels to

In the parallel-channel environment, performance and availability have been major factors in configuring DASD subsystems.

a storage control or to increase the number of storage controls per channel must be approached with analysis of the potential impact on performance, availability, and configuration management.

There are certain cases, such as installations with many LPARS (logical partitions defined by PR/SM) or systems images, where providing channel connectivity to many images is important. Installations with this requirement will need to review the performance considerations described below and configure appropriately. One point to consider is that the LPARS individually probably will not drive the DASD subsystem to the same extent that a single-image processor might, and thus some reduction in pathing may be satisfactory.

The effect of higher channel utilization should be somewhat less significant for a 3990 Model 3 than for a 3990 Model 2. Storage controls that exhibit channel utilizations in excess of 40 to 50 percent for extended periods probably should continue with the same number of ESCON paths as they currently have parallel paths. One consideration is that a reduction from four to three paths results in an asymmetric configuration, one cluster with two paths, the other with one. It is not clear that there is significant benefit to be gained in this reduction, since the impact of a cluster failure depends on which cluster failed—in one case, the storage control has two paths, and in the other, only one path. A reduction to two paths may be

possible, but performance in the event of a path failure may not be acceptable.

Another factor to think about when considering path reduction is the proportion of the workload on the storage control that is either dump-restore or intensive in sequential processing. Such applications are channel-intensive and may bypass cache. DFDSS (Data Facility Data Set Services) is one such program. Since many such applications run essentially at device speed, they may not be able to take advantage of the higher ESCON channel rates. Also, storage controls with lower hit ratios will be more sensitive to path reduction because of the increased likelihood of an RPS miss. This sensitivity is due to the fact that more operations are misses and work directly with the devices at device speed. 3990 Model 2 configurations may not be able to sustain good performance with fewer paths.

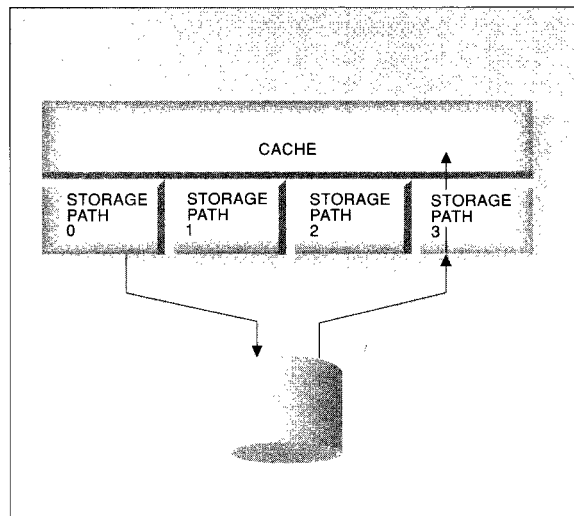
The same considerations apply to adding new storage controls to a channel.

Whenever possible, DLSE mode should be used. Even if there are only two channel paths from a given storage control to a system, the 3990 Model 3 is able to use DLSE internally for stage and destage operations. For example, if a stage request is initiated on storage path 1, the data may actually be transferred to cache via storage path 3 (Figure 11). Having a choice of four paths to complete the stage operation significantly reduces the impact of a storage path busy condition. Even a 3990 Model 2 can take advantage of DLSE mode: both storage paths within a multipath storage director would have to be busy before the 3990 would complete the I/O operation using the other cluster and a different controller within the DASD A-unit.

Migration to the ESCON environment

Asset protection, whether software or hardware, was a major consideration in the design of ESCON. The goal was to limit hardware upgrades or replacements and software modifications, as well as the effort required for hardware and software migration, procedural changes, training of the operations staff, and the disruption caused by the changes. By anticipating migration scenarios, we were able to provide for coexistence of the old and the new technologies.⁷

Figure 11 Use of DLSE for internal 3990 operations



Changes have been made to the system control programs, the channel subsystem, and the 3990 to support ESCON. We have already seen the changes brought about by the ECKD architecture and non-synchronous operation. To a very large extent, these changes are kept within the access methods and do not affect most application programs or the way in which data are stored on disk. Description of devices to the system is unchanged. Path management functions operate in the same way for ESCON as for parallel channels. IOCP has additional parameters to describe the new system components. The channel hardware, consisting of cables, connectors, channel adapters, and protocols, has been replaced. Within the 3990 Storage Control, the ESCON ports are new. The ESCON hardware and licensed internal code provide a new means of initializing the channel interfaces, and they implement the link and device protocols. Operation is nonsynchronous. The 3990 continues to accept the same channel programs that it always has. Description of an ESCON configuration in the vital product data has been simplified.

While this set of changes is indeed significant, it is important to note that the user and application interfaces are virtually untouched.

When introducing major architectural changes into a large computer system, a variety of techniques may be employed to make the transition

easier. We have already seen a couple of them: shielding a user's applications from the changes by using special routines that make the required modifications automatically whenever possible and providing independence from the transmission medium and protocols by embedding them in specialized hardware. Isolation of change in this way makes it easier to rely on existing capabilities and also easier to coexist with the older technology. Examples include path grouping, dynamic reconnect, and alternate path selection.

Maintaining as much as possible of the existing human interface also fosters smoother migration. Operator commands have not changed; we have added some new parameters and screens for the operator's use. Devices may still be addressed in the same fashion. It is particularly important where a device may be accessed using ESCON channels on one system and parallel channels on a sharing system. The IOCP process has not been replaced; it has been enhanced with new parameters, and a new hardware configuration dialog is available to provide interactive definition of a new configuration.

Provision for migration and coexistence between parallel and ESCON components. Large data processing installations usually have an extensive range of devices installed, both in terms of the type of equipment and in generations, such as a mix of 3990 and 3880 Storage Controls with 3380 and 3390 DASD. Migration plans for introducing the ESCON architecture must therefore accommodate a wide range of hardware capabilities and provide a great deal of flexibility in sequencing specific migration activities. Because of this, both the processors and the 3990 Storage Control support the intermix of ESCON and parallel channels. Furthermore, conversion devices may be used to accommodate either older DASD on processors equipped with ESCON channels or allow the attachment of 3990s with ESCON adapters to older processors.

The conversion devices provide some interesting configuration considerations and variations. There are two conversion devices: the IBM 9034 ESCON Converter Model 1 and the 9035 ESCON Converter Model 2.

The 9034 ESCON Converter Model 1. The 9034 ESCON Converter Model 1 permits attaching con-

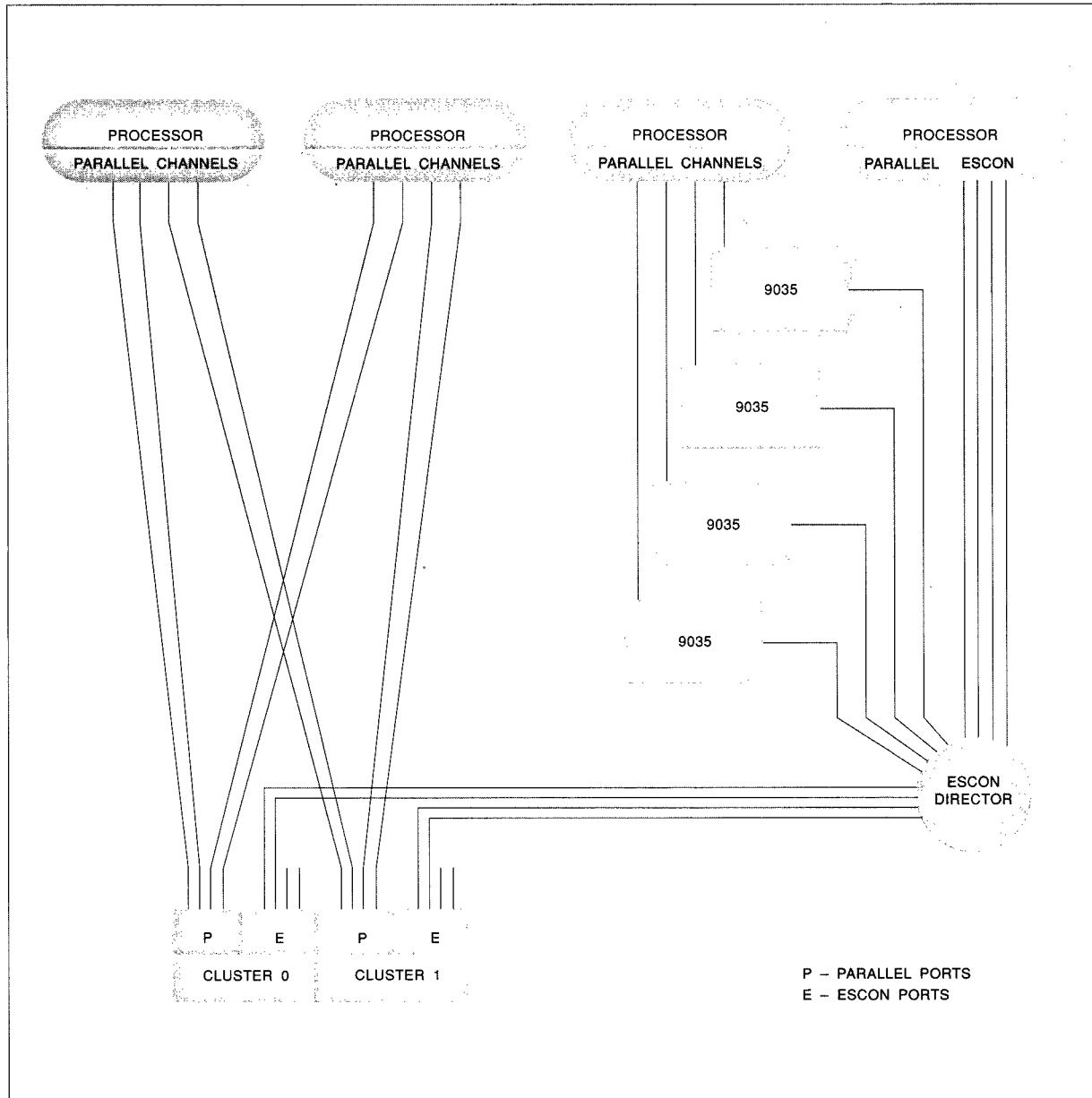
trol units with parallel interfaces (including the 3880) to ESCON channel hardware. The 9034 requires licensed internal code in the ESCON channel, specified in the IOCP, and loaded at the power-on-reset of the processor. It uses a special subset of the ESCON protocols that does not use the source and destination addresses essential to the operation of the ESCON Director. In particular, any director connections used with the 9034 must be static. As a result, it does not provide all of the ESCON advantages available when native ESCON channels are used with the ESCON Director.

However, the ESCON Director can be effectively used to provide similar switching capability to that provided by an IBM 3814 switch. In this case, the processors that need switching capability will all have fiber optic links installed from their ESCON channels to the director. The paths will be defined in the IOCP and SCP in a manner similar to the equivalent parallel configuration. Only one of the processors will have the channel paths on line and have a static connection through the director to the 3990 port. All of the other processors will have the corresponding paths off line and will not have static connections defined to the 3990 ports. Either the ESCON Director console or the ESCON Manager can be used to switch the static connection from one processor to another. Only one of the processors may be actively using the connection at one time; manual intervention is required to change it.

The 9034 does not provide the same extended distances that ESCON allows. For the 3990, the maximum distance supported is 1.2 kilometers, compared to the 9 kilometers supported in an ESCON configuration with two ESCON Directors. Furthermore, when an ESCON Director is inserted in the configuration to provide static switching capability, the maximum distance drops by 200 meters. A 3990 (or a 3880) running with a 9034 is still operating in synchronous mode, and so is still constrained to complete channel turnarounds within the DASD gap time.

The benefit of the 9034 is its ability to accommodate a parallel control unit. For the 3990 this ability is also its limitation: because these interfaces are parallel when attached to the 9034, and in fact use a multidropped connection where appropriate, the 3990 cannot take advantage of the dynamic change capabilities of ESCON. Removing or adding channel paths to such a 3990 will be po-

Figure 12 Typical configuration using 9035 ESCON Converters



tentially disruptive and will affect any other control units on the same channel paths.

Replacing parallel channels with ESCON channels configured for the 9034 has no effect on the attached control units; they still operate as parallel I/O control

units. All of the same rules for configuring parallel channels apply: up to eight control units on a channel, and a maximum parallel cable distance of 400 feet (including a deduction of 15 feet for each control unit attached). This distance is in addition to the maximum of 1.2 kilometers of fiber optic cable.²⁰

9035 ESCON Converter Model 2. The 9035 Converter, shown in Figure 12, is used to allow the attachment of a 3990 with ESCON adapters to a processor with parallel channels. Its principal use will be in an installation where one of the processors has been converted to ESCON, but at least one of the remaining processors is still parallel.

The 9035 relies on the 3990 for link-error reporting and for logical path establishment procedures. As a result, the 3990 contains special licensed internal code to support the ESCON functions that normally would have been handled by the ESCON channel link.

Because the 9035 uses the ESCON ports on a 3990, it operates in nonsynchronous mode, and therefore, the processor using the 9035 must support nonsynchronous DASD operations.

The 9035 can be considered an extension of the tailgate for the 3990. Each 9035 is associated with a unique port on the 3990. However, a 9035 and an ESCON channel may share the same 3990 port, as shown in Figure 12. In fact, sharing will probably be the most common configuration of a 9035. Similarly, two 9035s which are attached to different processor images may share a single 3990 port through an ESCON Director.

Summary

The products that comprise the ESCON family provide a rich set of functions with significant benefits for MVS-, VM-, and VSE-based systems. We have seen that these benefits arise from close interaction of many different components, ranging from the features of operating systems, to new hardware, to new microcode function, and to enhancements to existing products. Although some of these features enhance parallel channel operation as well, they provide the fullest benefit when used together in an ESCON configuration, again illustrating the strength of a system-wide design for new capabilities.

*Trademark or registered trademark of International Business Machines Corporation.

Cited references and note

1. Performance data contained in this document were determined in a controlled environment. Therefore, the results that may be obtained in other operating environ-

ments may vary significantly. Users should evaluate applicable data for each specific environment and may experience different results. **Consequently, this does not constitute a performance guarantee or warranty.**

2. *IBM 3990 Storage Control Reference*, GA32-0099, IBM Corporation; available through IBM branch offices.
3. *IBM Enterprise Systems Architecture/390 Principles of Operation*, SA22-7201, IBM Corporation; available through IBM branch offices.
4. *IBM System/360 and System/370 I/O Interface Channel to Control Unit OEMI*, GA22-6974, IBM Corporation; available through IBM branch offices.
5. *Introducing Enterprise Systems Connection*, GA23-0383, IBM Corporation; available through IBM branch offices.
6. C. P. Grossman, *IBM 3990 ESCON: Function, Installation, and Migration*, GG66-3213, IBM Corporation; available through IBM branch offices.
7. R. L. Clark, S. G. Glassen, C. P. Grossman, J. Mahan, K. R. Neslund, and D. S. Stary, *Enterprise Systems Connection: Planning for Migration*, GG66-3181, IBM Corporation; available through IBM branch offices.
8. *IBM 3990 Storage Control Introduction*, GA32-0098, IBM Corporation; available through IBM branch offices.
9. *Enterprise Systems Architecture/390 ESCON I/O Interface*, SA22-7202, IBM Corporation; available through IBM branch offices.
10. *Enterprise Systems/9000 and Enterprise Systems/3090 Input/Output Configuration Program Users Guide*, SC38-0097, IBM Corporation; available through IBM branch offices.
11. *MVS/ESA Hardware Configuration Definition—Using the Dialog*, GC33-6457, IBM Corporation; available through IBM branch offices.
12. *MVS/ESA Planning: Dynamic I/O Configuration*, GC28-1674, IBM Corporation; available through IBM branch offices.
13. *Introducing the Enterprise Systems Connection Manager*, GC23-0422, IBM Corporation; available through IBM branch offices.
14. *IBM 3990 Storage Control Planning, Installation, and Storage Administration*, GA32-0100, IBM Corporation; available through IBM branch offices.
15. *Functional Characteristics and Configuration Guide, Processor Complex Models 820 and 900*, GA22-7139, IBM Corporation; available through IBM branch offices.
16. M. Bohl, *Introduction to IBM Direct Access Storage Devices*, Science Research Associates, Chicago, IL (1981).
17. *Introduction to Nonsynchronous Direct Access Storage Systems*, GC26-4519, IBM Corporation; available through IBM branch offices.
18. C. P. Grossman, "Evolution of the DASD Storage Control," *IBM Systems Journal* 28, No. 2, 196-226 (1989).
19. C. P. Grossman, "Cache-DASD Storage Design for Improving System Performance," *IBM Systems Journal* 24, Nos. 3/4, 316-334 (1985).
20. *Introducing the 9034 Enterprise Systems Connection Converter*, GA23-0361, IBM Corporation; available through IBM branch offices.

General references

D. De Jong, *Extended Count Key Data and Nonsynchronous DASD I/O*, GG24-3571, IBM Corporation; available through IBM branch offices.

Enterprise Systems Architecture/390 Common I/O-Device Commands, SA22-7204, IBM Corporation; available through IBM branch offices.

Introducing the Enterprise Systems Connection Directors, GA23-0363, IBM Corporation; available through IBM branch offices.

Introducing the 9035 Enterprise Systems Connection Converter, GA23-0392, IBM Corporation; available through IBM branch offices.

MVS/ESA MVS Configuration Program Guide and Reference, GC28-1817, IBM Corporation; available through IBM branch offices.

Planning for the 9032 Enterprise Systems Connection Director, GA23-0364, IBM Corporation; available through IBM branch offices.

Planning for the 9034 Enterprise Systems Connection Converter, GA23-0362, IBM Corporation; available through IBM branch offices.

Accepted for publication October 11, 1991.

Carol P. Grossman *IBM Storage Systems Products Division, 9000 S. Rita Road, Tucson, Arizona 85706.* Mrs. Grossman is a senior planner. Her most recent activities have focused on the development and introduction of the ESCON adapter features of the IBM 3990 Storage Control. A major component of this activity has been in the area of implementation and migration planning for ESCON. She is the author of the IBM publication *IBM ESCON 3990: Function, Installation, and Migration* and coeditor of the IBM manual *ESCON Planning for Migration*. Other activities have included authorship of two papers on cache and storage controls in the *IBM Systems Journal*, as well as several other articles on similar subjects. She has presented a variety of 3990-related topics to U.S. GUIDE, European G.U.I.D.E., and SHARE for the last several years. She has been a planner for storage controls since the introduction of the IBM 3880 Model 13 in 1982. Before that, she was a storage systems specialist in Chicago, Illinois, where she joined IBM in 1974. Mrs. Grossman has a B.S. and an M.S. in mathematics from Northwestern University and is a member of Phi Beta Kappa and Pi Mu Epsilon.

Reprint Order No. G321-5466.